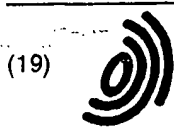


A 18



Europäisches Patentamt
European Patent Office
Office européen des brevets



(11) **EP 0 855 444 A2**

(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
29.07.1998 Bulletin 1998/31

(21) Application number: 98300573.7

(22) Date of filing: 27.01.1998

(51) Int Cl.⁶: **C12N 15/57, C12N 9/64,
C12N 1/21, C12N 5/10,
C07K 16/40, A61K 48/00,
A61K 39/395, A61K 31/70,
A61K 38/48, C12Q 1/68,
C12Q 1/37**

(84) Designated Contracting States:
**AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC
NL PT SE**

(30) Priority: 28.01.1997 GB 9701684

(71) Applicants:
• **SMITHKLINE BEECHAM PLC**
Brentford, Middlesex TW8 9EP (GB)
• **SMITHKLINE BEECHAM CORPORATION**
Philadelphia Pennsylvania 19103 (US)

(72) Inventors:
• **Powell, David J.,**
c/o SmithKline Beecham Pharm.
King of Prussia, PA 19460 (US)

• **Smith, Trudi S., c/o SmithKline Beecham Pharm.**
King of Prussia, PA 19460 (US)
• **Chapman, Conrad G.,**
c/o SmithKline Beecham Pharm.
Harlow, Essex CM19 5AW (GB)
• **Murphy, Kay SmithKline Beecham Pharm.**
Harlow Essex CM19 5AW (GB)

(74) Representative: **Crump, Julian Richard John et al**
1J Cleveland,
40-43 Chancery Lane
London WC2A 1JQ (GB)

(54) **Aspartic proteinase 2 (ASP2)**

(57) ASP2 polypeptides and polynucleotides and methods for producing such polypeptides by recombinant techniques are disclosed. Also disclosed are methods for utilizing ASP2 polypeptides and polynucle-

otides in the design of protocols for the treatment of Alzheimer's Disease, cancer, and prohormone processing, among others, and diagnostic assays for such conditions.

09 / 471, 669

EP 0 855 444 A2

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: +44 171 405 5875

(B) TELEFAX: +44 171 831 0749

(C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2541 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

25 ATGGCCCAAG CCCTGCCCTG GCTCCTGCTG TGGATGGGCG CGGGAGTGCT GCCTGCCCAC 60
 GGCACCCAGC ACGGCATCCG GCTGCCCTG CGCAGCGGCC TGGGGGGCGC CCCCCTGGGG 120
 CTGCGGCTGC CCCGGGAGAC CGACGAAGAG CCCGAGGAGC CCGGCCGGAG GGGCAGCTTT 180
 30 GTGGAGATGG TGGACAACCT GAGGGGCAAG TCGGGGCAGG GCTACTACGT GGAGATGACC 240
 GTGGGCAGCC CCCCAGAGAC GCTCAACATC CTGGTGGATA CAGGCAGCAG TAACTTTGCA 300
 GTGGGTGCTG CCCCCACCC CTTCCTGCAT CGTACTACC AGAGGCAGCT GTCCAGCACA 360
 TACCGGGACC TCCGGAAGGG TGTGTATGAG CCCTACACCC AGGGCAAGTG GGAAGGGGAG 420
 35 CTGGGCACCG ACCTGGTAAG CATCCCCCAT GGCCCCAAGC TCACTGTGCG TGCCAAACATT 480
 GCTGCCATCA CTGAATCAGA CAAGTTCTTC ATCAACGGCT CCAACTGGGA AGGCATCCTG 540
 GGGCTGGCCT ATGCTGAGAT TGCCAGGCCT GACGACTCCC TGAGACCTTT CTTTGACTCT 600
 CTGGTAAAGC AGACCCACGT TCCCAACCTC TTCTCCCTGC AGCTTTGTGG TGCTGGCTTC 660
 40 CCCCTCAACC AGTCTGAAGT GCTGGCCTCT GTCGGAGGGA GCATGATCAT TGGAGGTATC 720
 GACCACTCGC TGTACACAGG CAGTCTCTGG TATACACCCA TCCGGCGGGA GTGGTATTAT 780
 GAGGTGATCA TTGTGCGGGT GGAGATCAAT GGACAGGATC TGAAAATGGA CTGCAAGGAG 840
 TACAACATATG ACAAGAGCAT TGTGGACAGT GGCACCACCA ACCTTCGTTT GCCCAAGAAA 900
 45 GTGTTTGAAG CTGCAGTCAA ATCCATCAAG GCAGCCTCCT CCACGGAGAA GTTCCCTGAT 960
 GGTTCCTGGC TAGGAGAGCA GCTGGTGTGC TGGCAAGCAG GCACCACCCC TTGGAACATT 1020
 TTCCCAAGTCA TCTCACTCTA CCTAATGGGT GAGGTTACCA ACCAGTCCTT CCGCATCACC 1080
 ATCCTTCCGC AGCAATACCT GCGGCCAGTG GAAGATGTGG CCACGTCCCA AGACGACTGT 1140
 50 TACAAGTTTG CCATCTCACA GTCATCCACG GGCAGTGTTA TGGGAGCTGT TATCATGGAG 1200
 GGCTTCTACG TTGTCTTTGA TCGGGCCCGA AAACGAATTG GCTTTGCTGT CAGCGCTTGC 1260
 CATGTGCACG ATGAGTTCAG GACGGCAGCG GTGGAAGGCC CTTTGTGCAC CTGGACATG 1320
 55 GAAGACTGTG GCTACAACAT TCCACAGACA GATGAGTCAA CCCTCATGAC CATAGCCTAT 1380
 GTCATGGCTG CCATCTGCGC CCTCTTCATG CTGCCACTCT GCCTCATGGT GTGTCAGTGG 1440

GAG
is Glu
codon

CGCTGCCTCC GCTGCCTGCG CCAGCAGCAT GATGACTTTG CTGATGACAT CTCCCTGCTG 1500
 AAGTGAGGAG GCCCATGGGA GAAAGATAGA GATTCCTCTG GGACCACACC TCCGTGGTTC 1560
 5 ~~ACTTTGGTCA CAAGTAGGAG ACACAGATGG CACCTGTGGC CAGAGCACCT CAGGACCCCTC 1620~~
 CCCACCCACC AAATGCCTCT GCCTTGATGG AGAAGGAAAA GGCTGGCAAG GTGGGTTCCTA 1680
 GGGACTGTAC CTGTAGGAAA CAGAAAAGAG AAGAAAAGAAG CACTCTGCTG GCGGGAATAC 1740
 TCTTGCTCAC CTCAAATTTA AGTCGGGAAA TTCTGCTGCT TGAAACTTCA GCCCTGAACC 1800
 10 TTTGTCCACC ATTCCCTTAA ATTCTCCAAC CCAAAGTATT CTTCTTTTCT TAGTTTCAGA 1860
 AGTACTGGCA TCACACGCAG GTTACCTTGG CGTGTGTCCC TGTGGTACCC GGGCAGAGAA 1920
 GAGACCAAGC TTGTTTCCCT GCTGGCCAAA GTCAGTAGGA GAGGATGCAC AGTTTGCTAT 1980
 TTGCTTTAGA GACAGGGACT GTATAAACAA GCCTAACATT GGTGCAAAGA TTGCCTCTTG 2040
 15 AATTAATAAA AAAAAGTAGA TTGACTATTT ATACAAATGG GGGCGGCTGG AAAGAGGAGA 2100
 AGGAGAGGGA GTACAAAGAC AGGGAATAGT GGGATCAAAG CTAGGAAAGG CAGAAACACA 2160
 ACCACTCACC AGTCCTAGTT TTAGACCTCA TCTCCAAGAT AGCATCCCAT CTCAGAAGAT 2220
 GGGTGTGTGT TTCAATGTTT TCTTTTCTGT GGTGCAGCC TGACCAAAAG TGAGATGGGA 2280
 20 AGGGCTTATC TAGCCAAAGA GCTCTTTTTT AGCTCTCTTA AATGAAGTGC CCACTAAGGA 2340
 AGTTCCACTT GAACACATGG AATTTCTGCC ATATTAATTT CCATTGTCTC TATCTGGAAC 2400
 CACCCTTTAA TCTCTACATA TGATTAGGTC CAGCACTTGA AAATATTCCT AACCNNAATT 2460
 TGNCTTGGGG GCTTTGCNGN CCAGGTGCTA AAAGGGNTTG GGTAGGNGNC CNCTTNTATN 2520
 25 TNATNCCTNA AAAGGTTANN G 2541

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 501 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Ala Gln Ala Leu Pro Trp Leu Leu Leu Trp Met Gly Ala Gly Val
 1 5 10 15
 Leu Pro Ala His Gly Thr Gln His Gly Ile Arg Leu Pro Leu Arg Ser
 20 25 30
 Gly Leu Gly Gly Ala Pro Leu Gly Leu Arg Leu Pro Arg Glu Thr Asp
 35 40 45
 Glu Glu Pro Glu Glu Pro Gly Arg Arg Gly Ser Phe Val Glu Met Val
 50 55 60
 Asp Asn Leu Arg Gly Lys Ser Gly Gln Gly Tyr Tyr Val Glu Met Thr
 65 70 75 80
 Val Gly Ser Pro Pro Gln Thr Leu Asn Ile Leu Val Asp Thr Gly Ser

		85		90		95	
		Ser Asn Phe Ala Val Gly Ala Ala Pro His Pro Phe Leu His Arg Tyr					
5		100		105		110	
		Tyr Gln Arg Gln Leu Ser Ser Thr Tyr Arg Asp Leu Arg Lys Gly Val					
		115		120		125	
10		Tyr Glu Pro Tyr Thr Gln Gly Lys Trp Glu Gly Glu Leu Gly Thr Asp					
		130		135		140	
		Leu Val Ser Ile Pro His Gly Pro Asn Val Thr Val Arg Ala Asn Ile					
		145		150		155	160
15		Ala Ala Ile Thr Glu Ser Asp Lys Phe Phe Ile Asn Gly Ser Asn Trp					
		165		170		175	
		Glu Gly Ile Leu Gly Leu Ala Tyr Ala Glu Ile Ala Arg Pro Asp Asp					
		180		185		190	
20		Ser Leu Glu Pro Phe Phe Asp Ser Leu Val Lys Gln Thr His Val Pro					
		195		200		205	
		Asn Leu Phe Ser Leu Gln Leu Cys Gly Ala Gly Phe Pro Leu Asn Gln					
		210		215		220	
25		Ser Glu Val Leu Ala Ser Val Gly Gly Ser Met Ile Ile Gly Gly Ile					
		225		230		235	240
		Asp His Ser Leu Tyr Thr Gly Ser Leu Trp Tyr Thr Pro Ile Arg Arg					
		245		250		255	
30		Glu Trp Tyr Tyr Glu Val Ile Ile Val Arg Val Glu Ile Asn Gly Gln					
		260		265		270	
		Asp Leu Lys Met Asp Cys Lys Glu Tyr Asn Tyr Asp Lys Ser Ile Val					
		275		280		285	
35		Asp Ser Gly Thr Thr Asn Leu Arg Leu Pro Lys Lys Val Phe Glu Ala					
		290		295		300	
		Ala Val Lys Ser Ile Lys Ala Ala Ser Ser Thr Glu Lys Phe Pro Asp					
		305		310		315	320
40		Gly Phe Trp Leu Gly Glu Gln Leu Val Cys Trp Gln Ala Gly Thr Thr					
		325		330		335	
		Pro Trp Asn Ile Phe Pro Val Ile Ser Leu Tyr Leu Met Gly Glu Val					
		340		345		350	
45		Thr Asn Gln Ser Phe Arg Ile Thr Ile Leu Pro Gln Gln Tyr Leu Arg					
		355		360		365	
		Pro Val Glu Asp Val Ala Thr Ser Gln Asp Asp Cys Tyr Lys Phe Ala					
		370		375		380	
50		Ile Ser Gln Ser Ser Thr Gly Thr Val Met Gly Ala Val Ile Met Glu					
		385		390		395	400
		Gly Phe Tyr Val Val Phe Asp Arg Ala Arg Lys Arg Ile Gly Phe Ala					
		405		410		415	
55		Val Ser Ala Cys His Val His Asp Glu Phe Arg Thr Ala Ala Val Glu					
		420		425		430	

Same as in
Seq ID NO: 2

of 09/471,669

Gly Pro Phe Val Thr Leu Asp Met Glu Asp Cys Gly Tyr Asn Ile Pro
 435 440 445
 5 Gln Thr Asp Glu Ser Thr Leu Met Thr Ile Ala Tyr Val Met Ala Ala
 450 455 460
 Ile Cys Ala Leu Phe Met Leu Pro Leu Cys Leu Met Val Cys Gln Trp
 465 470 475 480
 10 Arg Cys Leu Arg Cys Leu Arg Gln Gln His Asp Asp Phe Ala Asp Asp
 485 490 495
 Ile Ser Leu Leu Lys
 500
 15

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2370 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GGCAGCTTTG TGGAGATGGT GGACAACCTG AGGGGCAAGT CGGGGCAGGG CTACTACGTG 60
 GAGATGACCG TGGGCAGCCC CCCGACAGCG CTCAACATCC TGGTGGATAC AGGCAGCAGT 120
 AACTTTGCAG TGGGTGCTGC CCCCCACCCC TTCTTGCAATC GCTACTACCA GAGGCAGCTG 180
 35 TCCAGCACAT ACCGGGACCT CCGGAAGGGT GTGTATGAGC CCTACACCCA GGGCAAGTGG 240
 GAAGGGGAGC TGGGCACCGA CCTGGTAAGC ATCCCCCATG GCCCCAACGT CACTGTGCGT 300
 GCCAACATTG CTGCCATCAC TGAATCAGAC AAGTTCTTCA TCAACGGCTC CAACTGGGAA 360
 GGCATCCTGG GGCTGGCCTA TGCTGAGATT GCCAGGCCTG ACGACTCCCT GGAGCCTTTC 420
 40 TTTGACTCTC TGGTAAAGCA GACCCACGTT CCCAACCTCT TCTCCCTGCA GCTTTGTGGT 480
 GCTGGCTTCC CCCTCAACCA GTCTGAAGTG CTGGCCTCTG TCGGAGGGAG CATGATCATT 540
 GGAGGTATCG ACCACTCGCT GTACACAGGC AGTCTCTGGT ATACACCCAT CCGGCGGGAG 600
 TGGTATTATG AGGTGATCAT TGTGCGGGTG GAGATCAATG GACAGGATCT GAAAATGGAC 660
 45 TGCAAGGAGT ACAACTATGA CAAGAGCATT GTGGACAGTG GCACCACCAA CCTTCGTTTG 720
 CCAAGAAAG TGTTTGAAGC TGCAGTCAAA TCCATCAAGG CAGCCTCTCC ACGGGAGAAG 780
 TTCCCTGATG GTTCTGGCT AGGAGAGCAG CTGGTGTGCT GGCAAGCAGG CACCACCCCT 840
 TGGAACATT TCCAGTCAT CTCACTCTAC CTAATGGGTG AGGTTACCAA CCAGTCCTTC 900
 50 CGCATCACCA TCCTTCCGCA GCAATACCTG CGGCCAGTGG AAGATGTGGC CACGTCCCAA 960
 GACGACTGTT ACAAGTTTGC CATCTCACAG TCATCCAC(3 GCACTGTTAT GGGAGCTGTT 1020
 ATCATGGAGG GCTTCTACGT TGTCTTTGAT CGGGCCCGAA AACGAATTGG CTTTGCTGTC 1080
 AGCGCTTGCC ATGTGCACGA TGAGTTCAGG ACGGCAGCGG TGGAAAGGCC TTTTGTCAAC 1140
 55 TTGGACATGG AAGACTGTGG CTACAACATT CCACAGACAG ATGAGTCAAC CCTCATGACC 1200

Xaa Leu Lys Arg Leu Xaa

770

5

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

10

(A) LENGTH: 19 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

15

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

20

GATGAGTTCA GGACGGCAG

19

(2) INFORMATION FOR SEQ ID NO:6:

25

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

30

(ii) MOLECULE TYPE: cDNA

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GGTGCCATAT GTGTCTCC

18

40

45 Claims

1. An isolated polynucleotide comprising a nucleotide sequence that has at least 80% identity over its entire length to a nucleotide sequence encoding the ASP2 polypeptide of SEQ ID NO:2; or a nucleotide sequence complementary to said isolated polynucleotide.

50

2. The polynucleotide of claim 1 wherein said polynucleotide comprises the nucleotide sequence contained in SEQ ID NO:1 encoding the ASP2 polypeptide of SEQ ID NO:2.

3. The polynucleotide of claim 1 wherein said polynucleotide comprises a nucleotide sequence that is at least 80% identical to that of SEQ ID NO: 1 over its entire length.

55

4. The polynucleotide of claim 3 which is polynucleotide of SEQ ID NO: 1.

5. The polynucleotide of claim 1 which is DNA or RNA.
6. A DNA or RNA molecule comprising an expression system, wherein said expression system is capable of producing a ASP2 polypeptide comprising an amino acid sequence, which has at least 80% identity with the polypeptide of SEQ ID NO:2 when said expression system is present in a compatible host cell.
7. A host cell comprising the expression system of claim 6.
8. A process for producing a ASP2 polypeptide comprising culturing a host of claim 7 under conditions sufficient for the production of said polypeptide and recovering the polypeptide from the culture.
9. A process for producing a cell which produces a ASP2 polypeptide thereof comprising transforming or transfecting a host cell with the expression system of claim 6 such that the host cell, under appropriate culture conditions, produces a ASP2 polypeptide.
10. A ASP2 polypeptide comprising an amino acid sequence which is at least 80% identical to the amino acid sequence of SEQ ID NO:2 over its entire length.
11. The polypeptide of claim 10 which comprises the amino acid sequence of SEQ ID NO:2.
12. An antibody immunospecific for the ASP2 polypeptide of claim 10.
13. A method for the treatment of a subject in need of enhanced activity or expression of ASP2 polypeptide of claim 10 comprising:
 - (a) administering to the subject a therapeutically effective amount of an agonist to said polypeptide; and/or
 - (b) providing to the subject an isolated polynucleotide comprising a nucleotide sequence that has at least 80% identity to a nucleotide sequence encoding the ASP2 polypeptide of SEQ ID NO:2 over its entire length; or a nucleotide sequence complementary to said nucleotide sequence in a form so as to effect production of said polypeptide activity *in vivo*.
14. A method for the treatment of a subject having need to inhibit activity or expression of ASP2 polypeptide of claim 10 comprising:
 - (a) administering to the subject a therapeutically effective amount of an antagonist to said polypeptide; and/or
 - (b) administering to the subject a nucleic acid molecule that inhibits the expression of the nucleotide sequence encoding said polypeptide; and/or
 - (c) administering to the subject a therapeutically effective amount of a polypeptide that competes with said polypeptide for its ligand, substrate, or receptor.
15. A process for diagnosing a disease or a susceptibility to a disease in a subject related to expression or activity of ASP2 polypeptide of claim 10 in a subject comprising:
 - (a) determining the presence or absence of a mutation in the nucleotide sequence encoding said ASP2 polypeptide in the genome of said subject; and/or
 - (b) analyzing for the presence or amount of the ASP2 polypeptide expression in a sample derived from said subject.
16. A method for identifying compounds which inhibit (antagonize) or agonize the ASP2 polypeptide of claim 10 which comprises:
 - (a) contacting a candidate compound with cells which express the ASP2 polypeptide (or cell membrane expressing ASP2 polypeptide) or respond to ASP2 polypeptide; and
 - (b) observing the binding, or stimulation or inhibition of a functional response; or comparing the ability of the cells (or cell membrane) which were contacted with the candidate compounds with the same cells which were not contacted for ASP2 polypeptide activity.
17. An agonist identified by the method of claim 16.